Location Based Services Analytics: Students as Co-Producers and Partners in Research

Richard J Self

University of Derby, UK

r.j.self@derby.ac.uk

Abstract. A perennial question in teaching is that of inspiring students to become innovative in their thinking and work. This paper presents one approach to developing excellence in undergraduate analytics research into a leading-edge topic which has had few publications to-date. The resultant analytics has provided valuable insights into the accuracy profiles of Location Services in smart devices and identified important consequences of the levels of accuracy that Assisted GPS (A-GPS) exhibits. Consolidation of the research of eight students by this author has resulted in approximately 3000 data points, from which it is possible to develop significant analytical insights which are of value to a wide range of potential users. It has also resulted in a raised academic and research profile for the University of Derby in the LBS field.

Keywords: analytics, location based services, accuracy profile, students as partners, A-GPS, SAS, SAS JMP, governance

1 Introduction

A perennial question in teaching is that of inspiring students to become innovative in their thinking and work. In addition, it is often difficult to get students to collect adequate quantities of data during their research projects in order that their analyses deliver reliable and statistically significant conclusions.

Research supervisors are generally familiar with the idea of gaining publications based on doctoral and post-doctoral research work. Few research supervisors, however, recognize the potential for coordinating teams of undergraduate students, in their final year research projects in leading-edge topics, to offer the opportunity for high quality conference papers and academic articles.

A recent final year dissertation / capstone project to characterize the accuracy of location services built into smart devices, such as smart phones and GPS cameras, demonstrates that it is possible to enthuse final year undergraduates to develop excellent research, collect large quantities of data, undertake good quality analysis, become co-authors of conference papers and gain valuable insights into a particular topic. The supervisor has been able to consolidate approximately 3000 data points into a single dataset for further analysis in order to gain greater validity and significance for analytical

technique development and further publications. It has also resulted in increased external reputation for the department and University within the business world.
This paper expands on the project, its operation and results, in order to encourage academics to explore their opportunities to both enthuse their students in research and to maximize their own opportunities for leading valuable research and publication opportunities.

2 The Project

2.1 Introduction

The author has made a habit of providing a single project specification each year on a topic for which there are relatively few recent publications, or which has not been previously researched. This is normally run on a two or three year cycle. The first year tends to be based on exploratory research to characterize the situation, following which the second year of the project cycle concentrates on opportunities and projects which are developed in order to provide diagnostic understanding of the insights developed during the first year of research.

2.2 Location Based Services Project

Location Based Services (LBS) in smart devices are based on the Assisted GPS (A-GPS) technology, which provides a faster acquisition of the initial location compared to standard GPS, as used in Sat Nav devices. GPS is claimed to provide an accuracy of 95% of locations to within ±10m. Many users have noticed that this level of accuracy is not delivered, however, there are very few academic analyses published that characterize the error functions of smart devices and the contexts of the error profiles. As a result, a project of this nature is clearly contributing to human knowledge and of significance to a range of users of the A-GPS functionality.
The **project context** was defined as:-

> There is a need to characterise the degree to which we can trust (if at all) the location data that is being captured by these various smart devices and whether different devices are more or less reliable than others. There are many consequences of these errors. Some are legal, some are related to organisational knowledge management, all impact the level of trust that we can have in recorded location data.

The **project research** was defined as:-

> Select a source of location data, such as from location tagged photos from a camera or smartphone, or the recorded data in a Sat Nav. Capture between 200 and 500 data points at known locations in a variety of circumstances (in-doors, in a vehicle, out-doors, etc. as relevant).
>
> For images, extract the EXIF data and calculate the positional error from the known positions where the images were collected and carry out an appropriate statistical analysis of the levels of uncertainty of the captured location data.

Potential **objectives** for the project were defined in relation to the **academic programmes** that students follow as:-

Forensics students

To investigate the scale of the problem and identify the impact on the Forensics Profession and the potential impact on "the Expert Witness" in court cases. This might include investigating the accuracy of location data contained in smart devices or Sat Navs.

Computer Science Students

To develop effective mechanisms to extract location data from suitable sources of location data, identify the magnitude of the location errors and to identify some of the contributory causes for the errors.

IT / Analytics Students

To collect a significant range of samples of location tagged data, to identify the location errors based on the known locations and to use SAS to perform the analysis and display the data in a suitable visualisation to create the greatest visual impact on the reader.

Computer Networks Students

A major source of location data that is relevant for Computer Networks professionals is the location of all the switches and routers in the internet. There are databases of the apparent locations of most of the switches in the world which can be visualised by tools such as Neotrace and many other tools. However, it is quickly apparent that many of the switches have incorrect or uncertain location data recorded. As a result it would be valuable to carry out the research necessary to characterise the level of accuracy and trustworthiness of this data, using appropriate statistical and visual approaches.

A total of 14 students joined the project, including 1 Forensics student and 1 Networks student, the remaining 12 were from the BSc IT programme (which is the University of Derby's Analytics programme).

2.3 Consolidated Analysis Requirement

It was made clear to all participants that a requirement of enrolling on the project was that their data would have to be consolidated into a single dataset for further analysis by the supervisor and that they would have to agree that this analysis of their data would be used as the basis for the supervisor's further academic publications. They were reassured that they would be given due attribution as the sources of the data and that there were opportunities for co-authorship of academic papers.

2.4 Research Ethics

Each student was required to complete a suitable Ethical Approval Form for their own research and the supervisor also gained Ethical Approval to consolidate and publish the resulting analyses.

3 Project Outcomes

3.1 Student Involvement

The students enthusiastically accepted the requirements for sharing their data and being able to gain publication attribution and acknowledgements. They were also very enthusiastic about the leading edge nature of the project and the fact that this would be the first comprehensive research into the LBS error issues.
A total of 2460 data points were collected by the eight contributing students, an additional 600 data points are in the process of collection and consolidation. The author is contributing a further 100 to 200 data points.

3.2 The Research Methodology

The students were required to carry out suitable literature research in order to develop the technical basis for the specific research questions for their own focus. This was the relatively easy part of their projects, due to the large amount of publications and sources relating to the technicalities of A-GPS. Each student was able to develop unique refined research questions which then developed into their experimental model and methodology.
As a result of the relative lack of prior publications characterizing the levels of accuracy of smart devices and A-GPS, the overall project had the nature of exploratory research in order to identify more refined research questions for future projects (Oates, 2006, p143)

3.3 Experiment Planning

A critical aspect for all of the students was to develop the means to identify as accurately as possible where each data point was collected and then to obtain the appropriate formula to calculate the distance between the measured coordinates and the true location.
All of the students used Google Maps as the basis to identify either a planned set of locations with known latitude and longitude, or found the location after collecting the data. Two of the students discovered an academic paper which characterizes the accuracy of Google Maps as a source of location data.
Students who used SAS to analyse the data used the GEODIST function, whereas those who used MS Excel, found a trigonometric function through Google. One of the latter students then used the known size of a tennis court to test the veracity of the trigonometric function. This latter student then analysed his data using SAS JMP, rather than Base SAS.
Several of the students used two smart phones, in order to identify differences in accuracy between their devices

3.4 Data Capture and Analysis

Each student obtained location fixes at between 150 and 200 locations, many used two devices, thus each contributed between 200 and 400 experimental data points. It is unlikely that any academic would have the time or patience to collect and locate 3000 data points! Locations were mainly in the UK, although one set has been collected in the Gulf States.
Students using MS Excel tended to use histograms as their primary analytical tool, whereas those using SAS were more likely to use Proc MEANS and Proc TTEST, as well as graphical presentations.

3.5 Project Insights

A range of interesting and valuable insights were developed, which will provide the focus for the project for next year. Among the most interesting were the following:-

- Eighty five percent of locations are accurate to within 25m, compared to the claimed 95% being accurate to 10m for GPS.
- Different generations of smart devices from a single manufacturer show statistically significant differences in accuracy.
- Smart devices from different manufacturers exhibit different levels of accuracy.
- The location coordinates at a specific location vary with time, often by more than 80m
- Weather conditions have an effect on the accuracy
- The environment has a significant impact, key among which are the Indoor / Outdoor / Rural / Urban / Residential contexts.
- LBS start-up locations (i.e. the first fix on starting the camera) can be very significantly in error, a 3km to 5km error is not unusual.

3.6 Project Impact

Overall the project identified that there are significant governance questions that need to be asked by those wishing to use LBS locations derived from smart devices using A-GPS. In particular, a very clear understanding must be developed of the accuracy requirements in relation to possible adverse consequences, such as reputational damage or actual physical danger for users of smart device navigational apps. The precise circumstances of each application need careful consideration together with the consequences of average errors of ±25m.
One student evaluated the Forensic and Criminological impact of the errors leading to the conclusion that neither prosecution nor defense teams can rely on LBS data extracted from smart devices without external corroborating evidence in criminal cases.

4 Future Research

The project insights provide strong guidance for the development of specific projects in the LBS field for the future, including the collection of more data relating to the time stability of the location fix, 3D analysis of the accuracy (including altitude) linked to the challenge of visualising the time-dependent 3D variability, including the creation of stereo pairs.

5 Academic Impact

5.1 Publications

Three of the students became co-authors of a paper that was presented at the **SAS Global Forum 2015** in Dallas, Texas, 26-29 April 2015, the other five were acknowledged for their contribution of data.
A total of eight students have been acknowledged in conference presentations at the **Location Based Services Conference**, London, UK, 2-3 June 2015 and the **University of Derby Annual Research Conference** in Derby, 3 July 2015.
A notable fact relating to the **LBS** conference in London was that the paper was the only one presented by a University academic, all the rest were presented by experts from industry.

5.2 Supervisory Impact

During the early stages of the project, whilst all the students were developing their literature reviews and methodology sections, all the students were seen together for an hour each week, as a single group. This enabled the students to both share literature resources and ideas and ensure that each project developed and maintained individuality and distinctiveness. This also led to the development of a wide range of unique insights.

6 Conclusion

This project has demonstrated that undergraduate students can become enthusiastic co-producers of world leading-edge research and partners in academic publications.

7 Bibliography

1. Oates, B, 2006, **Researching Information Systems and Computing**, Sage, London